Stock Share Prices Forecasting System using Case-Based Reasoning

Zarni Sann[1], Thi Thi Soe[2], Khaing Myat Nwe[3]
[1]*Faculty of Computer Systems and Technologies, University of Computer Studies (Mandalay), Myanmar.*
[2]*Faculty of Computer Science, University of Computer Studies (Mandalay), Myanmar .*
[3]*Faculty of Computer Systems and Technologies, University of Computer Studies (Mandalay), Myanmar.*

***Abstract*-- This work is to study how effective the Case-based Reasoning (CBR) method is to solve the forecasting stock share problem. Stock price forecasting is the act of determining the future value of a company stock on time series historical data. In real world, forecasting stock prices is not an easy task because this forecasting is based on variety of many financial conditions and it cannot be decided accurately. The system starts with inputting details of future stock share price. It is then matched with previous stored cases and retrieved the most similar cases. The local-global principle is applied to match cases and the similarity percentages are shown as a result. If the new attribute need to modify the previous case, the system can consider increasing the stock share with some financial factors. Finally, CBR generates a forecasted stock share result to increase or decrease. The results are to increase/decrease percentage amount of stock price.**

***Keywords*-- Case-based Reasoning (CBR), CBR inference engine, The local-global principle**

I. INTRODUCTION

CBR is a general paradigm for problem solving based on the recall and reuse of specific experiences. AI practitioners have applied and extended this paradigm for numerous classes of design problems, indicating that it is an important aspect of design support. CBR supports design by reminding designers of previous experiences that can help with new situation [1].

This system is based on historical data such as investment income, stock sale income, net income, and number of shares, total assets and total's liabilities of past time. In this system, Case-Based Reasoning (CBR) is used to forecast future stock share price. This system has used five years (2008, 2009, 2010, 2011, 2012) data that stored in database. Hence, storing and retrieving information on previous dataset are the most important tasks to provide business organization.

In this system, the features of stork share data are represented in attribute-value pairs. A similarity table is constructed for local similarity measure of each attribute. The values of the system user entered are considered as a new problem case. It is then matched with all the stored cases in historical database by using the local-global similarity measure. As a result, the system outputs the ranked similar cases. Based on the retrieved cases, the system user can modify or reuse it as the stock share price case solution. CBR is the artificial intelligent method that is suitable to solve the problem by finding similar cases from the past. Based on the similar case, the solution is to reuse the similar case and to revise its similar case solution [4, 5].

The paper aims at presenting a case-based reasoning system for forecasting stock share to support for business sector. The remainder of the paper is organized as follows. Section 2 presents the background theory, inference engine of CBR and the similarity measure utilized in this system. Section 3 illustrates the system design. Section 4 depicts the implementation of the system and conclusion is given in section 5.

1.1 Motivation and objectives of system

The motivation of the system is forecasting stock prices basically for financial profit gain. The aim of this system is to forecast stock price in Stock Exchange using Case-Base Reasoning for annual data from 2008 to 2012. Buying a stock share is the cost of purchasing on an exchange. Stock prices can be affected by a number of things including volatility in the market, current economic conditions, and popularity of the company. The main objectives of the system are: to predict the accurate forecasting result by using Knowledge base and four procedure, to gain the net profit from stock share forecasting, to support the business environment, and to determine stock share buying or not by using forecasting result.

II. BACKGROUND THEORY

Case-based reasoning (CBR) is a problem-solving paradigm that remembers previous similar situations (or cases) and reuses the information and knowledge about the stored cases for dealing with new problems. CBR systems have intuitive appeal because much of human problem solving capability is experience based, that is, humans draw on past experience when solving problems and can readily solve problems that are similar to ones encountered in the past.

On the other hand, the case-based reasoning is performed by the concept of parallel instead of serial chain like rule-base. The knowledge is stored in case like slots. The CBR works in cyclic form as the following procedures. When the user inputs

new problems to be solved, the system first of all searches the case base for the same case or similar cases. If the retrieved case meets with the user needs, reuse it as the current problem solution. Otherwise, modify the necessity parts to produce an optimal solution and output this case. At the same time, store the solution in case base to be capitalized and eventually reused as reference for a new problem [4, 5].

2.1 CBR inference engine

CBR module contains five parts: description of the issues, case retrieval, case assessment, case amendment, and case learning as shown in Fig. 1.

(1) Problem Description: User describes the problem to solve in functions, parameters and other elements.

(2) Case Retrieval: When initiating a case-based problem solving process, the first phase is the retrieval of useful cases providing solutions to be reused easily for solving the problem at hand. The quality of the case retrieval strongly depends on the quality of the used similarity measures.

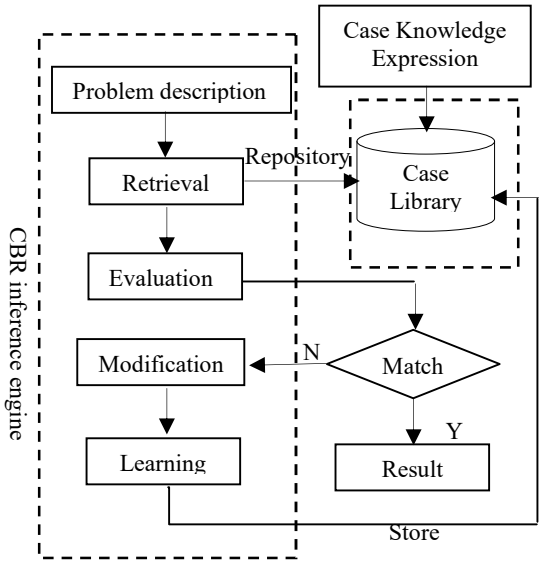

Fig. 1 CBR Inference Engine

(3) Case Assessment: The output of this module is the list of cases retrieved according to the ranked similarity of the matched case.

(4) Case Amendment: If instances are similar to the problem, but some aspects are still not suitable after assessment, the case amendment module can be used to amend the case, and create a new instance and store it.

(5) Case Learning: Methods of solving problems can be obtained from past experience. Case learning module can store the knowledge and experience gained in the previous steps [4, 5].

2.2 Similarity measure

A similarity measure is a function Sim:$D_D \times D_D \rightarrow [0,1]$. By computing the similarity between the query, q and the case characterizations of the cases contained in CB, the retrieval mechanism has to identify a list of cases, called retrieval result, ordered by the computed similarity values. The number of cases to be retrieved may be specified by an integer value which denotes the maximal number of cases to be retrieved [1].

2.2.1 The local-global principle

A similarity measure that can be adapted on a particular attribute-value based case representation is called the local-global principle. A local similarity measure for an attribute A is a function Sim_A: $A_{range} \times A_{range} \rightarrow [0,1]$, where A_{range} is the value range of A. The second important part of similarity measures defined according to the local-global principle is attribute weights. They are used to express the different importance of individual attributes for the entire utility approximation.

Here, we set an integer value 1 for less important attribute and 6 for more important one. The system allows setting these two values as the system user preferences.[1]

Global similarity measure is represented by an aggregation function computing the final similarity based on the local similarity values computed previously and the attribute weights defined:

A global similarity measure for D is a function

Sim: $D_D \times D_D \rightarrow [0, 1]$, of the following form:

$$\text{Case Sim (query, case)} = \frac{\sum_{i=1}^{n} w_i * sim_i}{\sum_{i=1}^{n} w_i}$$

where, sim_i is similarity of i^{th} feature, w_i is weight of i^{th} feature [1]

III. SYSTEM BACKGROUND AND DESIGN

System takes five years data set are storing in Case Base. Data sets include investment income, stock sales income, company's net income, number of shares, total assets, total liabilities and target output for stock share forecasting. The result are as type of class: increase(1) or decrease (0). Training for sample database as shown in table 1, and these data are more than 1000 times for year 2008.

- **Investment Income:** Income coming from interest payments, dividends, capital gains collected upon the sale of a security or other assets, and any other profit that is made through an investment vehicle of any kind.
- **Stock Sales Income:** Sale Income is income that a company receives from its normal business activities, usually from the sale of stock and services to customers.
- **Company's Net Income:** Net incomes an entity's income minus cost of goods sold, expenses and taxes for an

accounting period. It is computed as the residual of all revenues and gains over all expenses and losses for the period, and has also been defined as the net increase in stockholder's equity that results from a company's operations

- **Number of Shares:** Number of shares is the counts of shares in company.
- **Total Assets:** In financial accounting, an asset is an economic resource. Anything tangible or intangible that is capable of being owned or controlled to produce value and that is held to have positive economic value is considered an asset. Simply stated, assets represent value of ownership that can be converted into cash (although cash itself is also considered an asset)
- **Total Liabilities:** The aggregate of all debts an individual or company is liable for. Total liabilities can be easily calculated by summing all of one's short-term and long-term liabilities, along with any off balance sheet liabilities which corporations may incur [2, 3, 6].

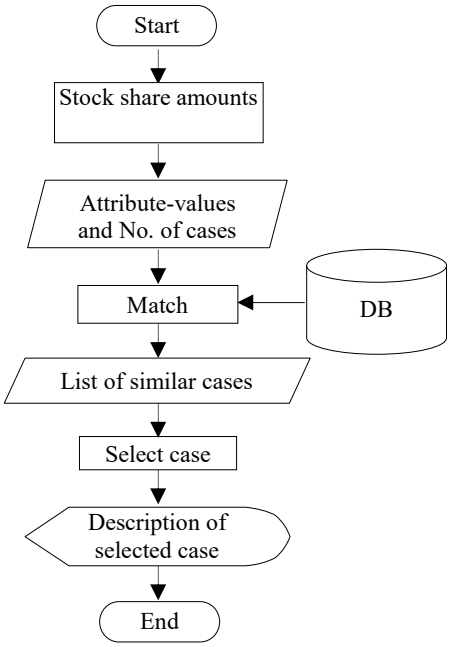

Fig. 2 System flow for case retrieval

For retrieval of useful cases for customer requirements, the matching of query case and the stored cases in the case base is performed by the similarity measure discussed in section 2.2. The computed similarity result is ranked and shown in percentage form. The user makes selection on one of the retrieved cases and views the detail description of the selected case. If the user meets with their needs, reuse the selected case as the current investment amount and increase the reuse times.

Otherwise, modify the necessity parts and re-input new attribute values. The system flow diagrams of retrieval and adaptation processes are illustrated in Fig. 2 and 3 respectively.

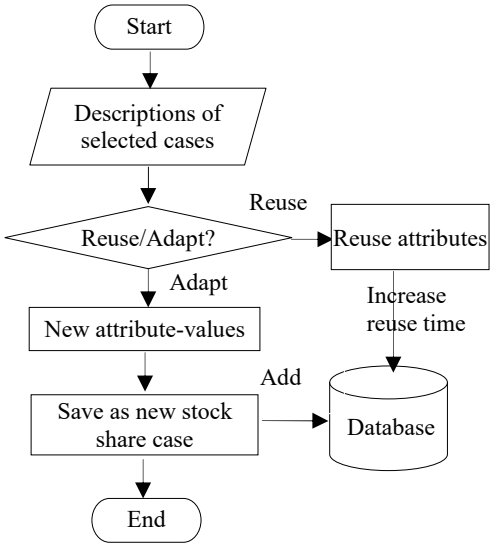

Fig. 3 System flow for case adaptation

As the case-base becomes grow, the searching time of the relevant cases gets slow. Therefore, we need to maintain the case-base. We will use the case addition and deletion policy. Case-addition option of case retain step is described in Fig. 3, in which the new stock share case is added to the database for future use.

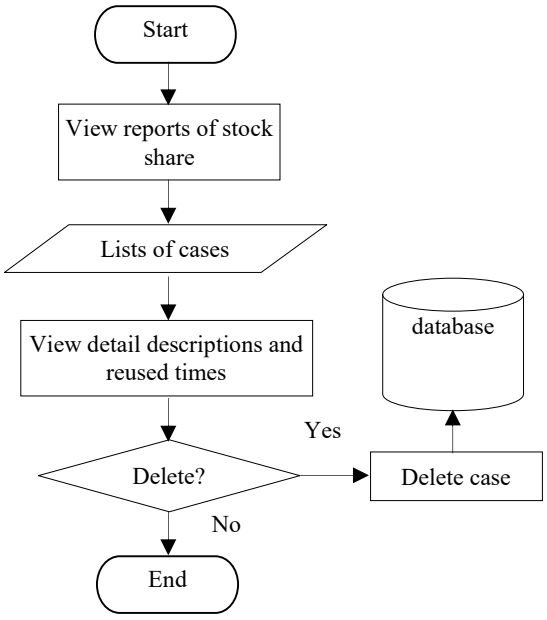

Fig. 4 System flow for case maintenance

However, if some stock share cases are not convenient for later use and the reuse time is low, we need to decide whether to delete it or not. The process of case-deletion is shown in Fig. 4.

TABLE I. SAMPLE DATABASE

Invest Income		Stock Sale Income	Net Income	No. of shares	Tot assets	Tot Liability	**Target Class**
Jan	2760	2810	2800	75	2800	510	1
Feb	2870	2900	2410	80	2410	620	0
Mar	2900	3030	2970	80	2970	550	1
April	2890	2990	2910	86	2910	500	1
May	2940	3020	2800	81	2800	610	1
June	3030	3060	3050	85	2050	630	0
July	3050	3100	3010	74	2040	580	0
Aug	3010	3030	3110	68	3110	600	1
Sept	3130	3080	3130	70	3130	610	0
Oct	3020	3100	2670	76	2670	700	1
Nov	3170	3230	2670	72	2670	710	1
Dec	3160	3110	2990	81	2990	720	1

IV. SYSTEM IMPLEMENTATION

This system used time series historical data (investment income, stock sale income, net income, number of shares, total assets, and total liabilities). In this implementation, the system will show the percentage of increase amount (1) or percentage of decrease amount (0). The system calculates the previous case with updated input values and then forecast of stock price may be increased or decreased.

A stock share price is the price of a single share of a number of saleable stocks of a company, derivative or other financial asset. The stock price or the value of a company is the present value of its future cash flows. The system forecasts the stock price situation that will increase or decrease with percentage amount. The system forecasts the stock price situation that will increase or decrease with percentage amount as shown in figure 5.

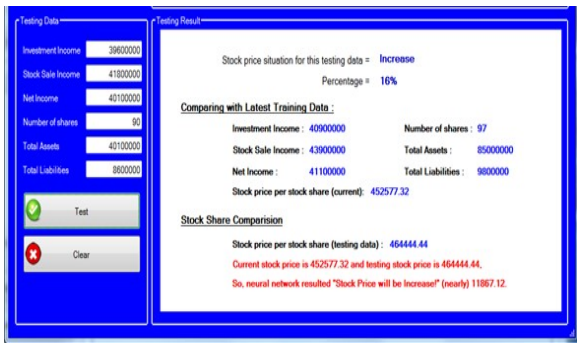

Fig. 5 Simulation of focusing result

V. CONCLUSION

Case-based reasoning concept is reviewed and applied in forecasting share price. A specific attributes case has better ability to improve the new attributes efficiency than clear rules. The reusable and adaptable ability of CBR is mainly utilized in this system. The work flow of the system is discussed with case study. As a result, the system user can get an exposure to the past remarkable attributes to produce new innovative share prices.

5.1 Advantages of the System

Stock market forecasting is the act of trying to determine the future value of a company stock or other financial instrument traded on a financial exchange. Predicting stocks is not an easy task. This system is developed as an application and was trained historical data to forecast a stock price in future. It can be implemented in any application.

5.2 Limitation and Further extension

Limitation of this system is five years dataset has used to predict the stock price and marking dataset is applied to forecast to stock share may be up or down. Stock Exchange Share Price is very hard to predict since there are no significant rules to estimate or predict that. Stock price prediction is one of the emerging fields of research and many methods like technical analysis, statistical analysis, time series analysis etc are used for this purpose.

REFERENCES

[1] Armin Stahl, "Learning of Knowledge-Intensive Similarity Measures in Case-Based Reasoning", PhD Dissertation, University of Kaiserslaitern, August, 2004.
[2] Atana Pal and Diptarka Chakraborty: Prediction of Stock Exchange Share Price using ANN and PSO. Department of Computer Science. Australia.
[3] Bruce Vanstone, Gavin Finnie, Tobias Hahn,, Stock market trading using fundamental variables and neural networks by 2010.
[4] Cunningham P, del Pobil AP, Mira J, Ali M. CBR: Strengths and weakness. In: Proc 11th Int Conf on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, Lecture Notes in Artificial Intelligence 1416, vol2; 1993, pp 517-523.
[5] Mary Lou Maher and Andres Gomez de Silva Garza, "Case-Based Reasoning in Design", AI in Design, IEEE, 1997, pp. 34-41.
[6] Ramon Lawrence, Using Neural Networks to Forecast Stock Market Prices, University of Manitoba, Department of Computer Science, 1997.